Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome

Beleut, Manfred ; Zimmermann, Philip ; Baudis, Michael ; Bruni, Nicole ; Bühlmann, Peter ; Laule, Oliver ; Luu, Van-Duc ; Gruissem, Wilhelm ; Schraml, Peter ; Moch, Holger

Abstract: ABSTRACT: BACKGROUND: Renal cell carcinoma (RCC) is characterized by a number of diverse molecular aberrations that differ among individuals. Recent approaches to molecularly classify RCC were based on clinical, pathological as well as on single molecular parameters. As a consequence, gene expression patterns reflecting the sum of genetic aberrations in individual tumors may not have been recognized. In an attempt to uncover such molecular features in RCC, we used a novel, unbiased and integrative approach. METHODS: We integrated gene expression data from 97 primary RCCs of different pathologic parameters, 15 RCC metastases as well as 34 cancer cell lines for two-way nonsupervised hierarchical clustering using gene groups suggested by the PANTHER Classification System. We depicted the genomic landscape of the resulted tumor groups by means of Single Nuclear Polymorphism (SNP) technology. Finally, the achieved results were immunohistochemically analyzed using a tissue microarray (TMA) composed of 254 RCC. Results: We found robust, genome wide expression signatures, which split RCC into three distinct molecular subgroups. These groups remained stable even if randomly selected gene sets were clustered. Notably, the pattern obtained from RCC cell lines was clearly distinguishable from that of primary tumors. SNP array analysis demonstrated differing frequencies of chromosomal copy number alterations among RCC subgroups. TMA analysis with group-specific markers showed a prognostic significance of the different groups. Conclusion: We propose the existence of characteristic and histologically independent genome-wide expression outputs in RCC with potential biological and clinical relevance.

DOI: https://doi.org/10.1186/1471-2407-12-310

BMC Cancer

BioMed Central
The Open Access Publisher

Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome

BMC Cancer 2012, **12**:310 doi:10.1186/1471-2407-12-310

Manfred Beleut (manfred.beleut@pareq.com})
Philip Zimmermann (phz@ethz.ch})
Michael Baudis (mbaudis@imls.uzh.ch})
Nicole Bruni (Nbruni@uhbs.ch})
Peter Bühlmann (buehlmann@stat.math.ethz.ch})
Oliver Laule (ola@nebion.com})
Van-Duc Luu (Vanducluu@yahoo.com})
Wilhelm Gruissem (wgruissem@ethz.ch})
Peter Schraml (Peter.Schraml@usz.ch})
Holger Moch (Holger.Moch@usz.ch})

ISSN 1471-2407

Article type Research article

Submission date 12 January 2012

Acceptance date 26 June 2012

Publication date 23 July 2012

Article URL http://www.biomedcentral.com/1471-2407/12/310

Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome

Manfred Beleut[1,5,*]
Email: manfred.beleut@pareq.com

Philip Zimmermann[2]
Email: phz@ethz.ch

Michael Baudis[3]
Email: mbaudis@imls.uzh.ch

Nicole Bruni[4]
Email: Nbruni@uhbs.ch

Peter Bühlmann[4]
Email: buehlmann@stat.math.ethz.ch

Oliver Laule[2]
Email: ola@nebion.com

Van-Duc Luu[1]
Email: Vanducluu@yahoo.com

Wilhelm Gruissem[2]
Email: wgruissem@ethz.ch

Peter Schraml[1*]
[*] Corresponding author
Email: Peter.Schraml@usz.ch

Holger Moch[1]
Email: Holger.Moch@usz.ch

[1] Institute of Surgical Pathology, University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich, Switzerland

[2] Department of Biology, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland

[3] Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[4] Seminar for Statistics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland

[5] PAREQ Research AG, Wagistrasse 14, 8952 Schlieren, Switzerland

* Corresponding author. PAREQ Research AG, Wagistrasse 14, 8952 Schlieren, Switzerland

Abstract

Background

Renal cell carcinoma (RCC) is characterized by a number of diverse molecular aberrations that differ among individuals. Recent approaches to molecularly classify RCC were based on clinical, pathological as well as on single molecular parameters. As a consequence, gene expression patterns reflecting the sum of genetic aberrations in individual tumors may not have been recognized. In an attempt to uncover such molecular features in RCC, we used a novel, unbiased and integrative approach.

Methods

We integrated gene expression data from 97 primary RCC of different pathologic parameters, 15 RCC metastases as well as 34 cancer cell lines for two-way nonsupervised hierarchical clustering using gene groups suggested by the PANTHER Classification System. We depicted the genomic landscape of the resulted tumor groups by means of Single Nuclear Polymorphism (SNP) technology. Finally, the achieved results were immunohistochemically analyzed using a tissue microarray (TMA) composed of 254 RCC.

Results

We found robust, genome wide expression signatures, which split RCC into three distinct molecular subgroups. These groups remained stable even if randomly selected gene sets were clustered. Notably, the pattern obtained from RCC cell lines was clearly distinguishable from that of primary tumors. SNP array analysis demonstrated differing frequencies of chromosomal copy number alterations among RCC subgroups. TMA analysis with group-specific markers showed a prognostic significance of the different groups.

Conclusion

We propose the existence of characteristic and histologically independent genome-wide expression outputs in RCC with potential biological and clinical relevance.

Keywords

DNA-microarray, SNP-array, RCC subgroups, Tissue microarray, Outcome

Background

Renal cell carcinoma (RCC) represents the most common malignancy arising in the adult kidney, with increasing incidence and poor prognosis [1]. RCC can be pathologically subdivided into different histological subtypes [2] based on the microscopic phenotype and the presence or absence of von Hippel-Lindau (*VHL*) gene alterations. The most frequent

histological subtype is clear cell RCC (ccRCC), followed by papillary RCC (pRCC) and chromophobe RCC (chRCC). Important prognostic parameters for RCC involve tumor and nodal stage [3].

In the search of critical genes, molecular studies identified several onco- and tumorsupressor gene candidates that are mutated and/or located within frequently gained and lost chromosomal regions of RCC [4-9]. Although multiple genes and signaling pathways have been implicated in renal cancer, *VHL* is the best characterized driver mutation, as it is mutated in the majority of sporadic ccRCC [10]. Loss of function of the VHL protein (pVHL) in ccRCC culminates in the deregulation of downstream target pathways that are important for uncontrolled cell proliferation and malignant progression [11]. In addition to *VHL,* alterations of the genes *MET*, *FH* and *BHD* are thought to be responsible for the development of familial RCC [2]. The low mutation frequencies reported for these genes in sporadic RCC subtypes [12-14], however, suggest other genes and pathways being relevant for the vast majority of RCC.

Microarray technology is an efficient approach to get comprehensive insights into individual and common tumor type-specific expression patterns based on hundreds of informative genes. Previous gene expression analyses using DNA microarrays suggested that unsupervised clustering combined with supervised learning methods optimize the molecular (re)classification of RCC to better predict cancer behavior. Distinct molecular expression profiles distinguishing between good and bad prognosis in RCC were identified [15-26]. However, the tumor samples were pre-selected according to histologic, clinical or molecular criteria in most of these studies. As a consequence, attempts to interpret general molecular strategies of RCC may have therefore been concealed by the co-appearance of surrogate markers. For example, although ccRCC is phenotypically and genotypically clearly different from pRCC and chRCC, we hypothesized that similar sets of common functional capabilities may exist in these tumor subtypes characterizable by the sum of molecular features occurring in RCC, irrespective of any histologic, clinical or single molecular parameters.

To test this hypothesis, we applied unsupervised clustering methods and integrated gene expression-, SNP and tissue microarray data using two independent sets of 146 and 254 RCC, respectively.

Methods

Renal cancer tissue, cell lines and nucleic acid extraction

Frozen primary RCC and tissue from RCC metastases were obtained from the tissue biobank of the University Hospital Zurich. This study was approved by the local commission of ethics (ref. number StV 38–2005). All tumors were reviewed by a pathologist specialized in uropathology (H.M.), graded according to the Fuhrman grading system and histologically classified according to the World Health Organization classification [2]. All tumor tissues were selected according to the histologically verified presence of at least 80 % tumor cells. Total RNA was extracted from 74 ccRCC, 19 pRCC, 2 chRCC, 2 mixed cc/pRCC and 15 metastases of ccRCC using the RNeasy minikit (Qiagen, Hilden, Germany). The quality of the RNA was measured using the Agilent Bioanalyzer 2100. DNA was extracted from 56 ccRCC, 13 pRCC and 69 matched normal renal tissues using the Blood and Tissue Kit (Qiagen). Expression analysis was additionally performed with RNA from 24 RCC cell lines,

6 cell lines from RCC metastasis and 4 prostate cancer cell lines as controls. All tumors and cell lines considered in expression and SNP-array experiments are listed in the supplementary data (see Additional file1: Table S1).

Microarrays and expression analysis

Reverse transcription of RNA, DNA labeling and hybridization on HG-U133A High-Throuput Arrays (Affymetrix, Santa Clara, USA) were performed at the Broad Institute of MIT and Harvard Medical School (Cambridge, MA, USA). Arrays were scanned using the HT Scanner. Affymetrix GeneChip data was normalized using MAS5 from Bioconductor [27] and $\log_2$-scaled. Hierarchical clustering was done with TIGR MeV [28] using Euclidian distance and average linkage. The identification of tumor type-specific biomarkers was performed using SAM [29]. The most significant genes were cross-checked in GENEVESTIGATOR [30] to remove probe sets that had absent calls across all samples. GENEVESTIGATOR is an online platform based on a high quality, manually curated database of microarray experiments enabling gene expression and regulation studies as well as the search for groups of genes sharing similar expression patterns by means of clustering and biclustering algorithms.

Generation and analysis of gene sets were performed with the PANTHER (Protein Analysis Through Evolutionary Relationships; http://www.pantherdb.org) Classification System database [31], by considering both, PubMed & Celera, datasets. The global functional overview of 17,181 human genes was extracted from PANTHER by using its standard settings. According to the developers, PANTHER is classifying all genes by their function through consideration of published scientific evidence and/or evolutionary relationships, therefore being able to even predict a function also in the absence of experimental evidence.

Affymetrix probe sets were identified for at least half of the genes extracted from PANTHER. For each gene set, a two-way hierarchical clustering of probe sets versus the complete set of expression arrays (146 arrays shown in Additional file1: Table S1) was run using GENEVESTIGATOR. We selected up to four clusters that best represented the overall array clustering in each pathway (see Additional file2: Figure S1 A-D). Finally, a joint clustering of all probe sets from these clusters resulted in the groupings described (Figure 1).

Figure 1 Molecular subclassification of renal cell carcinoma. Two-way hierarchical clustering of Affymetrix gene expression microarray data of 146 samples against the 92 pathway-related genes. Blue: relative increase-, white: relative decrease of gene expression. The PANTHER "pathway" affiliation of probe sets is indicated by colored barcode (right): green – "Inflammation"; pink – "Wnt"; orange – "Angiogenesis" and light blue – "Integrin" (see also Additional file3: Table S2). Note: None of the 4 groups is exclusively related to any of the "dominating pathways"

Statistical data validation of the three tumor groups

Random Forest and linear discriminant analysis were calculated using packages *random Forest* [32] and MASS *lda* [33] of R version 2.11.1, respectively. The testing of significance of variable selection was performed by group label shuffling; meaning by calculating the probability of finding the same classification accuracy for random groups with the same number of variables. Therefore, group labels were shuffled randomly 500 times, each time random forest was calculated, the variables were sorted by relevance and the 4 best variables

were then used for recalculation of random forest with same shuffled groups. Clustering was performed using function *hclust* R version 2.11.1.

SNP array analysis and classification

Labeling and hybridization of extracted DNA on Genome Wide Human SNP 6.0 arrays (Affymetrix) were performed at the Broad Institute of MIT and Harvard Medical School (Cambridge, MA, USA). Arrays were scanned using the GeneChip Scanner 3000 7 G. Raw probe data CEL files were processed with the R statistical software framework using the array analysis packages from the aroma.affymetrix project [34]. Total copy number estimates were generated using the CRMAv2 method [35] including allelic cross talk calibration, normalization for probe sequence effects and normalization for PCR fragment-length effects. Copy number segmentation was performed using the Circular Binary Segmentation method [36] implemented in the DNA copy package available through the Bioconductor project. Normalized data plots including segmentation results, oncogene map positions and known copy number variations as reported in the Database of Genomic Variants were generated with software packages developed for the Progenetix project [37]. Map positions were referenced with respect to the UCSC genome assembly hg18, based on the March 2006 human reference sequence (NCBI Build 36.1). Data from arrays with prominent probe level noise after normalization were excluded before proceeding with the evaluation of copy number imbalances. Overall, 114 SNP 6.0 arrays (55 RCC and 69 normal tissue samples) were used for final data processing.

For the generation of overall genomic imbalance profiles, probabilistic thresholds of 0.13/-0.13 were used for genomic gains and losses, respectively. Microarray and SNP data have been deposited in GEO under GSE19949.

Tissue microarray construction and immunohistochemistry

We used two TMAs with tumor tissue from 27 and 254 RCC specimens, respectively. The samples were retrieved from the archives of the Institute for Surgical Pathology; University Hospital Zurich (Zurich, Switzerland) between the years 1993 to 2003. In addition to tumor stage and Fuhrman grade, information about sarcomatoid differentiation was also available for all tumors. Areas with sarcomatoid differentiation were identified by one pathologist (H.M.) and defined as described [38].

TMAs were constructed as previously described [39]. To sufficiently address tumor heterogeneity, we used 3 punches per tumor for the construction of the TMA with 27 tumor samples. One biopsy cylinder per tumor was regarded as sufficient for constructing the TMA with 254 tumors. TMA sections (2.5 μm) on glass slides were subjected to immunohistochemical analysis according to the Ventana (Tucson, AZ, USA) automat protocols. CD34 (Serotec Ltd. - clone QBEND-10, dilution 1:800), MSH6 (BD Biosciences – clone 44, dilution 1:500) and DEK (BD Biosciences – clone 2, dilution 1:400) stainings were performed and analyzed under a Leitz Aristoplan microscope (Leica, Wetzlar, Germany). Tumors were considered MSH6 or DEK positive if more than 1 % of tumor cells showed unequivocal nuclear expression. MVD was determined as previously described [40]. Contingency table analysis and Pearson's chi-square tests were used to analyze the associations between protein expression patterns and clinical parameters. Overall survival rates were determined according to the Kaplan–Meier method and analyzed for statistical differences using a log rank test. A Cox proportional hazard analysis was used to test for

independent prognostic information. The statistics were performed with SPSS 18.0 for Windows (SPSS Inc., Chicago; IL)

Results

Gene expression patterns split RCC into three molecular groups

We chose PANTHER [31] to extract a standard overview of the classification of 17,181 human genes by their function. PANTHER allocates 6,017 of these genes into 145 "superior pathways". Four of these pathways involve more than 150 genes ("Wnt" 497 genes, "Inflammation" 476 genes, "Angiogenesis" 354 genes, "Integrin" 365 genes), others such as "Cysteine biosynthesis", listed only one gene.

We used the RNA extracted from 97 primary RCCs of different pathologic parameters, 15 RCC metastases and 34 cell lines (see Additional file1: Table S1), to identify any gene expression patterns in pathways containing more than 20 genes. For this purpose, we used the gene expression data obtained from Affymetrix HG-U133A arrays and performed two-way hierarchical clustering with each of those gene sets using GENEVESTIGATOR [30].

Only within the matrices of "Wnt", "Inflammation", "Angiogenesis", and "Integrin", which included process-related as well as downstream target genes as suggested by PANTHER, we observed clearly distinguishable major gene expression clusters. The most prominent gene expression clusters are highlighted in Additional file2: Figure S1 A-D and Additional file3: Table S2. Interestingly, no such differentiating gene expression patterns were obtained through hierarchical clustering of the genes of the remaining pathways (i.e. apoptosis or HIF-signaling) which, according to PANTHER, contained less than 150 genes (see Additional file2: Figure S1 E).

We next asked for possible relations among the different tumor groups and their specific gene expression patterns as detected from the 4 "dominating pathways". For this purpose, we selected with GENEVESTIGATOR up to four gene clusters from each of the four matrices encompassing a total of 92 genes, which were most representative for the overall clustering of the samples within each matrix (see Additional file2: Figure S1 A-D and Additional file3: Table S2) and combined them into a new matrix. Subsequent clustering of this matrix yielded four distinct groups (Figure 1). Notably, the 92 genes represented only a small percentage of genes involved in the suggested "dominating pathways" (see Additional file3: Table S2). Moreover, many of them (such as *MAPK*, *RHO*, *NOTCH*, *PDGF*, *RAS*, *JUN*, *ARF*, *PIK3*) also belong to other cancer-related pathways. Importantly, as none of the four groups was associated with any of those pathways (Figure 1 – color coding bar, right), we preferred to subdivide the groups into tumor groups "A", "B", "C" and "cell lines". Table 1 shows the 97 RCC specimens subdivided in groups A, B and C and characterized by tumor subtype, tumor stage and nuclear differentiation grade. Most interestingly, primary RCC split into group A, B or C, irrespective of their clinical characteristics (see also Additional file1: Table S1).

Table 1 Classification of two RCC sets and their clinical characteristics

		RCC microarray set				RCC TMA set			
		A	B	C		A	B	C	
		N (%)	N (%)	N (%)	N (total)	N (%)	N (%)	N (%)	N (total)
Histological	ccRCC	48 (65)	16 (22)	10 (13)	74	39 (27)	66 (45)	41 (28)	146
subtype	pRCC	1 (5)	6 (32)	12 (63)	19	0	17 (52)	16(48)	33
	chRCC	0	1 (50)	1 (50)	2	0	7 (70)	3 (30)	10
	cc/pRCC	0	0	2 (100)	2	-	-	-	-
Tumor stage	pT1/pT2	32 (52)	16 (26)	14 (22)	62	27 (28)	48 (51)	20 (21)	95
	pT3/pT4	17 (49)	7 (20)	11 (31)	35	12 (14)	39 (45)	36 (41)	87
Fuhrman grade	grade 1	3 (43)	1 (14)	3 (43)	7	1 (100)	0	0	1
	grade 2	27 (63)	8 (19)	8 (19)	43	20 (36)	20 (36)	15 (27)	55
	grade 3	18 (44)	12 (29)	11 (27)	41	16 (20)	44 (55)	20 (25)	80
	grade 4	1 (17)	2 (33)	3 (50)	6	2 (4)	24 (47)	25 (49)	51
sarcomatoid	yes	nd	nd	nd		3 (7)	20 (43)	23 (50)	46
	no	nd	nd	nd		36 (26)	67 (48)	37 (26)	140

nd: not done due to limited tissue material

Gene delineation for stable RCC stratification

To confirm that the expression status of our 4 groups is specific, we profiled gene expression across 40 primary RCC samples arbitrarily chosen from the three RCC groups. Five independent hierarchical clusterings of these samples across arbitrarily chosen and pathway independent probe sets as well as a clustering against all 22,000 probe sets of the Affymetrix array showed that group B was clearly distinct from A and C. Notably, group A always appeared as a tight cluster within the C clad (Figure 2A left and Additional file4: Figure S2). These findings confirmed the previous subgrouping of RCC based on the selected 92 genes (Figure 1) and moreover suggests the presence of genome wide, discrete and group-specific gene expression signatures.

Figure 2 Genome-wide expression signatures in RCC. A. Hierarchical clustering of 40 RCC samples across all probe sets of the HG-U133A array, identifying the 3 groups (left). Hierarchical clustering of the 40 RCC samples based on expression signal values from 769 genes identified from the SNP array analysis, show diffuse clusters prior to group acquaintance (middle), but are unraveling the 3 RCC groups when individual tumors are affiliated (here: color coded) to their respective group before clustering (right). **B – C.** Heatmaps of RCC group-specific signatures with corresponding intensity bars (absolute values). Relative increase (yellow) and relative decrease (blue) of gene expression. **B**. Gene expression of the 50 best classifiers of subgroup B against subgroups A/C across a subset of A, B and C RCC. **C**. Gene expression of the 24 best classifiers of subgroup A against subgroup C across a subset of A and C RCC subgroups

Identification of the best RCC group identifiers

By using SAM [29], at least a 2-fold change in the expression level was seen for more than 2,000 genes, with 1,455 genes being higher and 715 genes being lower expressed in group B compared to A/C, and 221 genes positively and 11 genes negatively regulated in A versus C.

The most differentially regulated genes between group B and groups A/C were represented by 48 genes, with 16 low expressed in B but strongly expressed in A/C (8.7 – 5.7 fold change) and 32 transcripts abundant in B but decreased in A/C (14.4 – 5.2 fold change) (Figure 2B; Additional file5: Table S3). Twenty-three genes clearly distinguished groups A and C with 4 genes highly expressed in C but not in A (14.3 – 2.5 fold change), while 19 were highly expressed in A but not in C (16.0 – 4.2 fold change) (Figure 2C; Additional file6: Table S4).

Statistical significance of the three RCC groups

The groups "A", "B" and "C" were further investigated towards accuracy and reproducibility of their classification. Using Random Forest, classification accuracy reached 96.94 % if only four variables out of the >22,000 measured genes were selected. The three groups separated very well as only 3 of the 97 measurements were misclassified under these conditions (Figure 3A). To test whether these three groups are outstanding, label shuffling of the groups and retrying classification with the four best variables was performed. Shuffled groups were analyzed for how often the same or a better classification accuracy than the original was achieved. Label shuffling considered that clusters A, B and C each contained 49 individuals and most "subtypes" in cluster A were ccRCC. Therefore at least one third of A was occupied with randomly selected ccRCC from B and C. 500 times label shuffling and classification trials resulted in zero times same or better classification accuracy ($p < 0.002$).

Figure 3 Validation and prognostic significance of the genome-wide expression signatures in RCC. A. Linear discriminant analysis of groups "A", "B" and "C" with 4 selected variables (genes). Classification of the three groups using the 4 highest ranked variables of Random Forest allows linear discriminant analysis (LDA) with 96.94 % accuracy. **B.** Kaplan–Meier analysis of tumor-specific survival in 176 RCC patients. Subgroup A (high MVD, DEK and MSH positive), B (high or low MVD, MSH6 negative) and C (low MVD, DEK and MSH positive) (log rank test: $p < 0.0001$)

Integration of DNA copy number alterations (CNAs) to the three RCC subgroups

In a first step, we analyzed genomic profiles of 45 RCC and corresponding normal tissues using Affymetrix 6.0 SNP arrays. We extracted an overall summary of detected genomic imbalances using Progenetix [37] and compared them to the entire available dataset of 568 RCC in the Progenetix database at censoring time (see Additional file7: Figure S3A). Consistent with previous CGH data [41], our results confirmed the overall composite of CGH profiles in RCC.

In order to clarify whether our three RCC groups are characterized by combinations and/or frequencies of specific CNAs, we analyzed our groups using CNA data from 36 RCC for which high quality SNP- and gene expression microarray data were available and allocated

and color coded with regard to Figure 1 (see also Additional file1: Table S1), 20 tumors in group A, 3 in group B and 13 in group C (see Additional file8: Table S5). By displaying all CNAs mapped to 811 cytogenetic bands (UCSC - hg18 cytogenetic mapping; chromosomes 1–22), all chromosomes were affected including the known RCC subtype-specific genomic alterations 3p-, 5q + (ccRCC) and 7+, 17+, 20+ (pRCC). This result is in line with previous CGH data [41]. Notably, loss of 3p was observed in all 3 groups and increased genomic derangements were seen in groups B and C compared to group A (see Additional file7: Figure S3 B).

Next, in order to identify the genes residing in minimally affected CNAs that possibly also directly contribute to the group-specific output signatures, we focused on tumor-specific genomic changes below 5 Mb which is the approximate resolution limit for chromosomal losses and gains obtained by chromosomal CGH [42]. We found 126 different regions in our cohort varying between 0.5 kb to 5 Mb and encompassing 61 allelic gains and 65 allelic losses (see Additional file9: Table S6). These chromosomal regions harbored coding regions of a total of 769 genes. Interestingly, in contrast to large chromosomal aberrations commonly detected by CGH in public data sets, the genomic alterations <5 Mb could not be linked to morphologically defined RCC subtypes. By looking at all chromosomal changes occurring in our RCC set, we found a unique cytogenetic "fingerprint" characteristic for each tumor. Despite this uniqueness we were able to allocate all RCC to one of the 3 groups at the gene expression level (Figure 1).

Unsupervised hierarchical clustering of the 769 CNA-affected genes (see Additional file9: Table S6) against the 40 arbitrarily selected primary RCCs (see chapter "Gene delineation") showed rather diffuse RCC clusters (Figure 2A middle) indicating at first no direct linkage to the three RCC signatures. However, as it was already demonstrated with randomly, CNA non-affected, picked gene sets (see Additional file4: Figure S2), the 769 CNA-affected genes could eventually be assigned to the three RCC groups, but only by knowing the three specific groups before clustering (Figure 2A right).

Molecular RCC grouping is an independent, survival-associated prognostic factor

We finally asked whether RCC of the three groups could also be classified by characteristic morphologies or specific expression patterns on the protein level. For this purpose, we randomly selected 9 RCC from each of the three respective groups (Figure 1) and placed them into a small tissue microarray (TMA). A Hematoxylin/Eosin stained TMA section was blindly evaluated by a pathologist (H.M.). All nine tumors of group A were characterized by high microvessel density (MVD), whereas there were no specific morphologic features in the tumors of groups B and C. To further verify this finding, we immunohistochemically stained the endothelial cell marker CD34 in the 27 RCC. As shown in Additional file10: Table S7, the results largely confirmed group-specific angiogenic traits. All nine tumors in group A, but only three in group B and one in group C had more than 100 microvessels, whereas the remaining ones had less than 50 microvessels per arrayed spot (0.036 mm^2). Tumors with high and low MVD were classified accordingly. In order to find more group-specific markers which separate group B tumors from A/C and group A tumors from C in combination with the CD34 staining, we further searched genome wide with SAM [29] for genes with a clear present or absent expression profile in the three groups. SAM identified several candidates, including *DEK* and *MSH6*, for which well-established antibodies were available. By examining immunostaining patterns of several protein candidates coded by these genes, we

were able to assign tumors with high MVD as well as DEK and MSH6 positivity to group A, high or low MVD and MSH6 negative tumors to group B, and tumors with low MVD but DEK and MSH6 positivity to group C. Examples of immunostained RCC are shown in Additional file11: Figure S4.

To evaluate the obtained group-specific protein expression patterns in a much higher number of tumors, we screened a TMA with 254 RCC. By strictly applying the staining combinations obtained from the small test TMA, 189 tumors (75 %) were clearly assigned to a specific group. The pathologic characteristics of the tumors assigned to the three RCC groups are shown in Table 1. There were organ-confined and metastatic RCC of different tumor subtype and nuclear differentiation grade with varying frequencies in these groups. To determine the clinical aggressiveness of these groups, we focused our analysis on 176 of 189 RCC samples on the TMA for which survival data were available. Kaplan-Meier analysis showed a highly significant correlation (log rank test: $p < 0.0001$) of group affiliation with overall survival, in which patient outcome was best in group A and worst in group C (Figure 3B). This result was independent from tumor stage and grade in a multivariate analysis (see Additional file12: Table S8). By performing this survival analysis, we demonstrate that the molecular re-classification of RCC allows the identification of early stage tumors (pT1 and pT2) with high metastasizing potential associated with poor patient prognosis. In addition, the finding of late stage RCC in group A also suggests the existence of patients with a relative good prognosis although their tumors were categorized as pT3.

Discussion

In this study we used unsupervised hierarchical clustering and gene expression pattern combination approaches to detect robust molecular clusters which classify RCC into three molecular groups with distinct prognostic values. In many previous studies, RCC cases were either preselected or expression data were linked according to pathologic and clinical criteria for further analysis [15-26]. Potential markers may have therefore represented surrogate traits, overall confirming phenotypes at a molecular basis. To our knowledge, patterns of gene expressions, independent of pathological or single molecular parameters pointing to general RCC biology remained uncovered to date.

Classification system databases suitable for comprehensive gene expression clustering

To identify common RCC gene expression signatures, we searched for large gene sets using the classification systems INGENUITY (http://www.ingenuity.com/), KEGG (http://www.genome.jp/kegg/) and PANTHER (http://www.pantherdb.org). Our gene expression analyses demonstrated that more than 150 genes are required to obtain major and clearly distinguishable gene clusters. In contrast to Ingenuity and KEGG, only PANTHER is able to integrate several hundred genes into “superior pathways”. Only within the clustering of these four dominating processes different major group patterns were obtained. The number of genes in the remaining pathways was too low and therefore not suitable for cluster analysis. It is important to understand that it was not our intention to analyze specific pathways within RCC. We rather used this platform to visualize sets of gene expression clusters which are differentially regulated within different RCC. There was no notable association between any of the RCC groups and any of the 4 pathways. The 92 genes extracted from the 4 matrices (pathways) were rather equally distributed over the 4 RCC

groups suggesting the partial involvement of all 4 pathways in the RCC groups. In our opinion the results of the clustering by using randomly selected 5 sets of about 700 genes clearly indicate that we could have taken any arbitrary chosen gene list for clustering independent of any pathways.

Unsupervised versus supervised clustering

For supervised analysis of gene expression patterns in tumors algorithms are commonly used that are linked to known clinical parameters such as tumor subtype, metastatic-non metastatic or treated-untreated. Consequently, the number of clusters to be expected is already known. As we tried to identify non-biased gene expression patterns, we chose unsupervised analysis for which the resulting numbers of clusters are unknown. To circumvent this problem we combined the strongest gene expression patterns into a new matrix and re-clustered them by using the second clustering step, importantly, against the same tumor cohort (Additional file 2: Figure S1 A-D and Figure 1). Our approach to randomly select genes and re-cluster them demonstrated that the three tumor groups remained stable (Additional file 4: Figure S2). We, therefore, believe that our two-times-two-way non-supervised clustering method is an alternative strategy to re-classify tumor types independent of TNM criteria. We cannot rule out that additional groups exist which may appear if more samples are included in the analysis.

Molecular signatures strictly separate RCC tissue from RCC cell lines

Surprisingly, the expression signature yielded from the renal cancer cell lines was clearly distinguishable from those derived from renal cancer tissues. We observed that individual cell line expression profiles, independent of their respective primary tumors, were all similar to each other. This general finding may mainly be caused by culture conditions, the artificial environment and the two-dimensional structure of cell culture layers. We therefore believe that expression profiling using cell lines would never lead to the detection of common renal cancer tissue-specific signatures. This also raises concerns about the possibility of discovering novel strategies for diagnosis and therapies by using *in vitro* systems only.

Molecular signatures do not coincide with pathologic criteria

In contrast to the cell lines which represent a separate group, RCC metastases and primary RCC split into group A, B or C, irrespective of the tumor subtype, stage, differentiation grade or sarcomatoid differentiation. When looking at RCC group A, which contains almost only ccRCC, it seems that the clustering results correlate with the histological subtype. However, these ccRCC were of different tumor stage and grade. The same is true for the tumors in group B and C. In these groups cRCC, pRCC as well as chRCC of different pathologic parameters were allocated. Furthermore, our molecular classification allows to additionally refine the staging and grading of tumors. Organ-confined RCC, particularly pT1 tumors, generally considered to have a good prognosis can further be subdivided in group A (good), B (worse) or C (worst) which also may have predictive impacts. Although ccRCC, pRCC and chRCC have a different morphological background, the combined appearance of the three histological subtypes across different clusters suggests molecular and functional similarities.

The three RCC output signatures are not influenced by the VHL/HIF axis

Based on the results obtained from a series of previous *VHL* mutation analyses, it is widely accepted that the loss of function of pVHL mainly contributes to the development of ccRCC [43]. The inactivation of pVHL leads to HIF-α stabilization and, hence, to the upregulation of a number of genes involved in RCC progression (i.e. *VEGFA, PDGFB, TGFA, CXCR4, CA9*) [44-46]. Therefore, we assumed to detect gene expression patterns connected to HIF signaling pathways. However, gene expression patterns demonstrated no remarkable linkage between HIF-regulated pathways and any of the RCC subgroups. This finding is in line with the results of a recent study in which *VHL* wild-type tumors, HIF-1α and HIF-2α overexpressing tumors, as well as HIF-2α-only overexpressing tumors were found in both ccRCC clusters [26].

We also looked at the *VHL* mutation status in all analyzed ccRCC and identified gene sequence alterations in the majority of the tumors [10]. A recent study demonstrated that the thermodynamic stability and the functionality of pVHL is dependent of the location and the type of mutation [10]. As the frequencies and types of *VHL* mutations were similar in all three RCC groups it was not surprising that there was no association with the gene expression patterns, neither with the *VHL* mutation status nor with any HIF-driven pathways (data not shown). Our data strongly suggest the existence of pVHL-independent mechanisms, resulting in distinct gene expression outputs which reflect common biologic pathways in renal cell cancer.

RCC gene expression signatures are not directly linked to copy number alterations

Our integrative approach that combined SNP- and microarray data, revealed no direct correlation between the signatures of CNA-affected genes analyzed in 45 RCC and the three RCC groups. Only one of the 92 cluster forming genes (*ITGAL*; see Additional file 3: Table S2) belonged to the 769 genes residing within the 126 CNAs found in our RCC set. Moreover, hierarchical clustering of both CNA-affected and non-affected genes demonstrated that the three RCC gene expression patterns are not directly influenced by copy number alterations. This finding is in line with a recent study which also found many discrepancies between CNA and gene expression [47]. The authors suggest that the expression of many "driver" genes are less correlated with their copy number than "passenger" genes due to selective pressure. Additional multiple ways exist to up- or down-regulate a gene.

RCC is not caused by alteration of single genes and pathways

It is remarkable that, although type and frequencies of CNAs were largely differing within the tumor cohort and varied between none (!) and 18 altered genomic regions in single tumors, each of the three group-specific gene expression patterns remained stable. We postulate that each of these RCC must have developed individual mechanisms in addition to CNAs (i.e. mutations, methylations, transcriptional and translational modifications), which together support the regulation of molecular components to reach one of the three tumor groups. A recent study showing that low CNA rates in tumors are related to increased levels of global DNA methylation and *vice versa* [48] supports our hypothesis.

In contrast to previous approaches, we combined several subtypes of RCC for non-supervised hierarchical clustering approaches in combination with LDA entirely unbiased from different clinico-pathologic parameters. Our results demonstrate that RCC group formation patterns remained very similar across various sets of genes arguing for a substantial number of genes which participate in the molecular definition of a RCC group. It is therefore not surprising that more than one third of the human genes have already been identified as cancer-relevant [49] and many of them being claimed as potential biomarkers [50]. As a consequence, we believe that in a tumor many molecular pathways must be directly or indirectly affected to eventually reach one of the three output signatures.

Characterization of the three RCC groups at the protein level

By subsequently performing our TMA analysis on a second, larger cohort of RCC we validated our results also on the protein level. To find appropriate markers we tested several antibodies directed against proteins whose genes were clearly upregulated in one of the groups. Among 10 candidates tested only MSH6, a DNA mismatch repair enzyme, and DEK, a chromatin- and RNA-associated protein mutated or overexpressed in certain cancers, showed reliable immunostaining results. The third protein, CD34, was indirectly identified by retrospectively analyzing the tumors histologically *after* the clustering analyses (Figure 1). We found increased microvessel density in group A by selecting the RCC samples randomly without knowing any specific pathological features (with the exception of stage and grade). Although not expressed in RCC cells, this endothelial marker is an ideal marker to morphologically distinguish group A from group B and C. Our effort to select suitable protein markers for the RCC groups demonstrated strong differences between the expression signatures at the RNA and the protein levels. Further protein analyses are needed to identify additional markers or marker combinations with both prognostic and predictive value.

Conclusion

We believe that the identified genome-wide signatures point to common molecular programs characteristic for the biology of RCC. Here, we provide a novel concept for RCC classification implying potential impacts on tumor diagnostics and the development of tailor-made therapies. We still do not know whether the identified signatures are restricted to RCC or exist also in other cancer types. If the latter is true these expression patterns may represent outputs of molecular events which have led to common functional characteristics of cancers.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

M.Be. and P.S. designed the study, analyzed, combined and interpreted data, and drafted the manuscript. P.Z., O.L. and W.G. provided technical support, processed and analyzed expression array data. M.Ba. provided technical support, processed raw SNP data and prepared Gene Expression Omnibus (GEO) information. N.B. & P.B. performed statistical significance analyses. VDL prepared expression array analyses and extracted nucleic acids.

H.M. & W.G. designed the study and H.M. reviewed all tumors. All authors read and approved the final manuscript.

Acknowledgements

We thank the Broad Institute Center for Genotyping and Analysis of MIT & Harvard, MA 02142 Cambridge, for performing SNP and expression array experiments; Walter J. Storkus and Kirsten Mertz for SLR cell lines. We are indebted to Martina Storz, Susanne Dettwiler and Silvia Behnke for outstanding technical assistance and Dieter Zimmermann and Gunther Boysen for valuable suggestions and critical manuscript reading. The project was supported by the Swiss National Science Foundation (3238BO-103145) to H.M., the Zurich Cancer League to H.M. and the Swiss Initiative in Systems Biology (SystemsX) to H.M., P.Z. and W.G.

References

1. McLaughlin JK, Lipworth L, Tarone RE: **Epidemiologic aspects of renal cell carcinoma.** *Semin Oncol* 2006, **33(5):**527–533.

2. Eble JN, Sauter G, Epstein JI, Sesterhenn IA: *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*. Lyon: IARC Press; 2004.

3. Moch H, Artibani W, Delahunt B, Ficarra V, Knuechel R, Montorsi F, Patard JJ, Stief CG, Sulser T, Wild PJ: **Reassessing the current UICC/AJCC TNM staging for renal cell carcinoma.** *Eur Urol* 2009, **56(4):**636–643.

4. Latif F, Tory K, Gnarra J, Yao M, Duh FM, Orcutt ML, Stackhouse T, Kuzmin I, Modi W, Geil L, *et al*: **Identification of the von Hippel-Lindau disease tumor suppressor gene.** *Science* 1993, **260(5112):**1317–1320.

5. Beroukhim R, Brunet JP, Di Napoli A, Mertz KD, Seeley A, Pires MM, Linhart D, Worrell RA, Moch H, Rubin MA, *et al*: **Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney.** *Cancer Res* 2009, **69(11):**4674–4681.

6. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, *et al*: **Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes.** *Nature* 2010, **463(7279):**360–363.

7. Morrissey C, Martinez A, Zatyka M, Agathanggelou A, Honorio S, Astuti D, Morgan NV, Moch H, Richards FM, Kishida T, *et al*: **Epigenetic inactivation of the RASSF1A 3p21.3 tumor suppressor gene in both clear cell and papillary renal cell carcinoma.** *Cancer Res* 2001, **61(19):**7277–7281.

8. Schmidt L, Duh FM, Chen F, Kishida T, Glenn G, Choyke P, Scherer SW, Zhuang Z, Lubensky I, Dean M, *et al*: **Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas.** *Nat Genet* 1997, **16(1):**68–73.

9. Contractor H, Zariwala M, Bugert P, Zeisler J, Kovacs G: **Mutation of the p53 tumour suppressor gene occurs preferentially in the chromophobe type of renal cell tumour.** *J Pathol* 1997, **181(2):**136–139.

10. Rechsteiner MP, von Teichman A, Nowicka A, Sulser T, Schraml P, Moch H: **VHL gene mutations and their effects on hypoxia inducible factor HIFalpha: identification of potential driver and passenger mutations.** *Cancer Res* 2011, **71(16):**5500–5511.

11. Frew IJ, Krek W: **pVHL: a multipurpose adaptor protein.** *Sci Signal* 2008, **1(24):**pe30.

12. Schmidt L, Junker K, Nakaigawa N, Kinjerski T, Weirich G, Miller M, Lubensky I, Neumann HP, Brauch H, Decker J, *et al*: **Novel mutations of the MET proto-oncogene in papillary renal carcinomas.** *Oncogene* 1999, **18(14):**2343–2350.

13. Morris MR, Maina E, Morgan NV, Gentle D, Astuti D, Moch H, Kishida T, Yao M, Schraml P, Richards FM, *et al*: **Molecular genetic analysis of FIH-1, FH, and SDHB candidate tumour suppressor genes in renal cell carcinoma.** *J Clin Pathol* 2004, **57(7):**706–711.

14. Khoo SK, Kahnoski K, Sugimura J, Petillo D, Chen J, Shockley K, Ludlow J, Knapp R, Giraud S, Richard S, *et al*: **Inactivation of BHD in sporadic renal tumors.** *Cancer Res* 2003, **63(15):**4583–4587.

15. Boer JM, Huber WK, Sultmann H, Wilmer F, von Heydebreck A, Haas S, Korn B, Gunawan B, Vente A, Fuzesi L, *et al*: **Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array.** *Genome Res* 2001, **11(11):**1861–1870.

16. Jones J, Otu H, Spentzos D, Kolia S, Inan M, Beecken WD, Fellbaum C, Gu X, Joseph M, Pantuck AJ, *et al*: **Gene signatures of progression and metastasis in renal cell cancer.** *Clin Cancer Res* 2005, **11(16):**5730–5739.

17. Kosari F, Parker AS, Kube DM, Lohse CM, Leibovich BC, Blute ML, Cheville JC, Vasmatzis G: **Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness.** *Clin Cancer Res* 2005, **11(14):**5128–5139.

18. Lenburg ME, Liou LS, Gerry NP, Frampton GM, Cohen HT, Christman MF: **Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data.** *BMC Cancer* 2003, **3:**31.

19. Schuetz AN, Yin-Goen Q, Amin MB, Moreno CS, Cohen C, Hornsby CD, Yang WL, Petros JA, Issa MM, Pattaras JG, *et al*: **Molecular classification of renal tumors by gene expression profiling.** *J Mol Diagn* 2005, **7(2):**206–218.

20. Skubitz KM, Skubitz AP: **Differential gene expression in renal cell cancer.** *J Lab Clin Meds* 2002, **140(1):**52–64.

21. Sultmann H, von Heydebreck A, Huber W, Kuner R, Buness A, Vogt M, Gunawan B, Vingron M, Fuzesi L, Poustka A: **Gene expression in kidney cancer is associated with**

cytogenetic abnormalities, metastasis formation, and patient survival. *Clin Cancer Res* 2005, **11(2 Pt 1):**646–655.

22. Takahashi M, Rhodes DR, Furge KA, Kanayama H, Kagawa S, Haab BB, Teh BT: **Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification.** *Proc Natl Acad Sci U S A* 2001, **98(17):**9754–9759.

23. Takahashi M, Yang XJ, Sugimura J, Backdahl J, Tretiakova M, Qian CN, Gray SG, Knapp R, Anema J, Kahnoski R, *et al*: **Molecular subclassification of kidney tumors and the discovery of new diagnostic markers.** *Oncogene* 2003, **22(43):**6810–6818.

24. Vasselli JR, Shih JH, Iyengar SR, Maranchie J, Riss J, Worrell R, Torres-Cabala C, Tabios R, Mariotti A, Stearman R, *et al*: **Predicting survival in patients with metastatic kidney cancer by gene-expression profiling in the primary tumor.** *Proc Natl Acad Sci U S A* 2003, **100(12):**6958–6963.

25. Young AN, Amin MB, Moreno CS, Lim SD, Cohen C, Petros JA, Marshall FF, Neish AS: **Expression profiling of renal epithelial neoplasms: a method for tumor classification and discovery of diagnostic molecular markers.** *Am J Pathol* 2001, **158(5):**1639–1651.

26. Brannon AR, Reddy A, Seiler M, Arreola A, Moore DT, Pruthi RS, Wallen EM, Nielsen ME, Liu H, Nathanson KL, *et al*: **Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns.** *Genes Cancer* 2010, **1(2):**152–163.

27. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, *et al*: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5(10):**R80.

28. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J: **TM4 microarray software suite.** *Methods Enzymol* 2006, **411:**134–193.

29. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98(9):**5116–5121.

30. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P: **Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes.** *Adv Bioinformatics* 2008, **2008:**420747.

31. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD: **PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium.** *Nucleic Acids Res* 2010, **38(Database issue):**D204–D210.

32. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP: **Random forest: a classification and regression tool for compound classification and QSAR modeling.** *J Chem Inf Comput Sci* 2003, **43(6):**1947–1958.

33. Venables WN, Ripley BD: *Modern Applied Statistics with S*. 4th edition. New York: Springer; 2002.

34. Bengtsson H, Wirapati P, Speed TP: **A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6.** *Bioinformatics* 2009, **25(17):**2149–2156.

35. Bengtsson H, Ray A, Spellman P, Speed TP: **A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods.** *Bioinformatics* 2009, **25(7):**861–867.

36. Venkatraman ES, Olshen AB: **A faster circular binary segmentation algorithm for the analysis of array CGH data.** *Bioinformatics* 2007, **23(6):**657–663.

37. Baudis M, Cleary ML: **Progenetix.net: an online repository for molecular cytogenetic aberration data.** *Bioinformatics* 2001, **17(12):**1228–1229.

38. de Peralta-Venturina M, Moch H, Amin M, Tamboli P, Hailemariam S, Mihatsch M, Javidan J, Stricker H, Ro JY, Amin MB: **Sarcomatoid differentiation in renal cell carcinoma: a study of 101 cases.** *Am J Surg Pathol* 2001, **25(3):**275–284.

39. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4(7):**844–847.

40. Schraml P, Struckmann K, Hatz F, Sonnet S, Kully C, Gasser T, Sauter G, Mihatsch MJ, Moch H: **VHL mutations and their correlation with tumour cell proliferation, microvessel density, and patient prognosis in clear cell renal cell carcinoma.** *J Pathol* 2002, **196(2):**186–193.

41. Baudis M: **Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data.** *BMC Cancer* 2007, **7:**226.

42. Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, Gooden GC, Ethier SP, Kallioniemi A, Kallioniemi OP: **Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data.** *Cancer Res* 2000, **60(16):**4519–4525.

43. Gossage L, Eisen T: **Alterations in VHL as potential biomarkers in renal-cell carcinoma.** *Nat Rev Clin Oncol* 2010, **7(5):**277–288.

44. Brugarolas J: **Renal-cell carcinoma–molecular pathways and therapies.** *N Engl J Med* 2007, **356(2):**185–187.

45. Grabmaier K, AdW MC, Verhaegh GW, Schalken JA, Oosterwijk E: **Strict regulation of CAIX(G250/MN) by HIF-1alpha in clear cell renal cell carcinoma.** *Oncogene* 2004, **23(33):**5624–5631.

46. Staller P, Sulitkova J, Lisztwan J, Moch H, Oakeley EJ, Krek W: **Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL.** *Nature* 2003, **425(6955):**307–311.

47. Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, Causton HC, Pochanard P, Mozes E, Garraway LA, Pe'er D: **An integrated approach to uncover drivers of cancer.** *Cell* 2010, **143(6):**1005–1017.

48. Poage GM, Christensen BC, Houseman EA, McClean MD, Wiencke JK, Posner MR, Clark JR, Nelson HH, Marsit CJ, Kelsey KT: **Genetic and epigenetic somatic alterations in head and neck squamous cell carcinomas are globally coordinated but not locally targeted.** *PLoS One* 2010, **5(3):**e9651.

49. Huret JL, Senon S, Bernheim A, Dessen P: **An Atlas on genes and chromosomes in oncology and haematology.** *Cell Mol Biol (Noisy-le-Grand)* 2004, **50(7):**805–807.

50. Poste G: **Bring on the biomarkers.** *Nature* 2011, **469(7329):**156–157.

Additional files

Additional_file_1 as XLS

Additional file 1: Table S1. List of samples used in expression array, 55 of them were also used for SNP experiment.

Additional_file_2 as PDF

Additional file 2: Figure S1. The strategy to find group-specific expression signatures in RCC. Hierarchical clustering of HG-U133A microarray probe sets representing genes from the Angiogenesis **(A)**, Inflammation **(B)**, Integrin **(C)**, and Wnt **(D)** "pathways" as annotated by PANTHER, across a set of 146 microarrays from our RCC experiment. For each "pathway", up to four probe set clusters (red boxes) were selected and combined for subsequent re-clustering. **(E)** Another PANTHER "pathway" (Apoptosis) and one RCC-relevant "pathway" (HIF). Note the presence of less genes in these matrices compared to A-D and the absence of clear probe set clusters (except for cell lines in "Apoptosis", indicated by the green bottom line), visually subdividing the matrix.

Additional_file_3 as XLS

Additional file 3: Table S2. List of clusters and containing genes, picked from separate "pathway clusterings" to be combined into one matrix.

Additional_file_4 as PDF

Additional file 4: Figure S2. The three RCC gene expression signatures spread genome-wide. Hierarchical clustering of 5 times arbitrarily chosen probe sets, each composed of ca. 660 genes) against group affiliated tumors (individual group-sample is labeled as A_, B_ or C_) **(A-E).** Note the tumor-group forming coincidence within the 5 independent analyses and the similarity with that shown in Figure 2A.

Additional_file_5 as XLS

Additional file 5: Table S3. List of top 48 genes with expression values, specific for RCC tumors of group B, relative to A and C.

Additional_file_6 as XLS

Additional file 6 Table S4. List of top 23 genes with expression value, distinguishing RCC tumors of group A from goup C.

Additional_file_7 as PDF

Additional file 7: Figure S3. The landscape of CNAs in RCC does not correlate with novel molecular subgroups. **(A**) Regional genomic CNAs in RCC shown as percentage of analyzed cases (genomic gains: yellow, up; losses: blue, down). Top: depiction of the overall CNAs in the 45 study cases; Down: published chromosomal and array CGH RCC data accessible through the Progenetix database (568 cases). Copy number variants (CNVs) were not filtered from the study case data besides application of a 100 kb size limit. Note the similar profiles. **(B)** Case specific regional copy number imbalances in 36 RCC study cases with regional genomic gain or loss status matched to 811 cytogenetic regions. The genomic profiles are randomly arranged within their subtypes. White areas indicate concurrent gain and loss in this cytoband. Note the appearance of known subtype-specific genomic alterations (3p deletions, 5q gains identifying clear cell RCC – asterisk and arrow/left side; gains of chromosomes 7, 17 and 20 identifying papillary RCC - arrows right side).

Additional_file_8 as XLS

Additional file 8 Table S5. List of 36 RCC tumors considered on expression- and SNP array, and their affiliation to a specific group according to gene expression array.

Additional_file_9 as XLS

Additional file 9: Table S6. List of tumorspecific regions (0–5 Mb) and involved genes, identified by SNP experiment.

Additional_file_10 as XLS

Additional file 10: Table S7. The Test Tissue Microarray to establish antibody combinations for tumor/group affiliations.

Additional_file_11 as PDF

Additional file 11: Figure S4. Examples of immunostained RCC group-specific markers CD34, DEK and MSH6. ccRCC with CD34-stained vascular microvessels (A, B); ccRCC with strong nuclear DEK (C) and MSH6 (D) positivity.

Additional_file_12 as XLS

Additional file 12: Table S8. Cox proportional hazard regression analysis for survival.

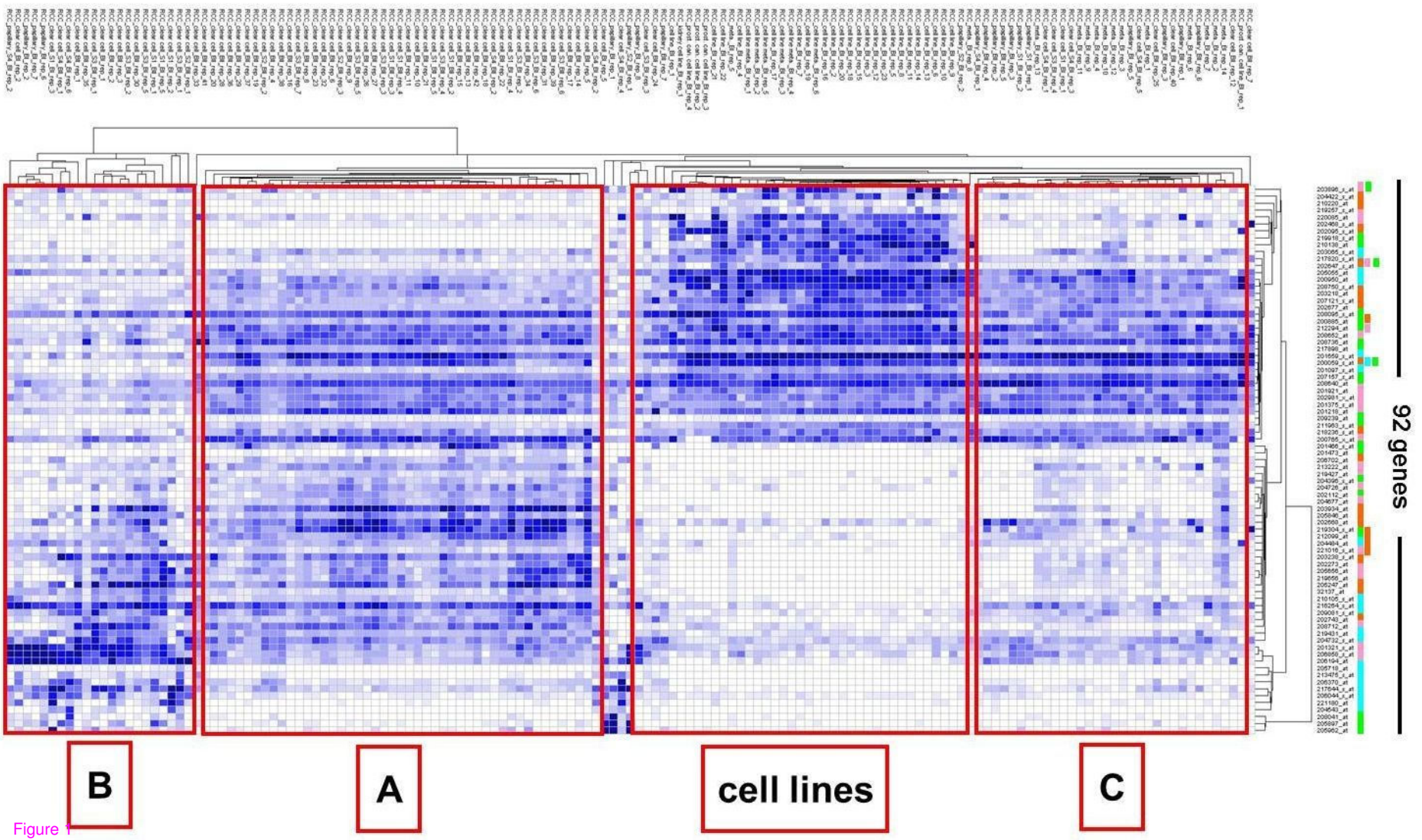
97 renal tumors, 15 metastases , 34 cell lines
92 genes
B
A
cell lines
C

Figure 1

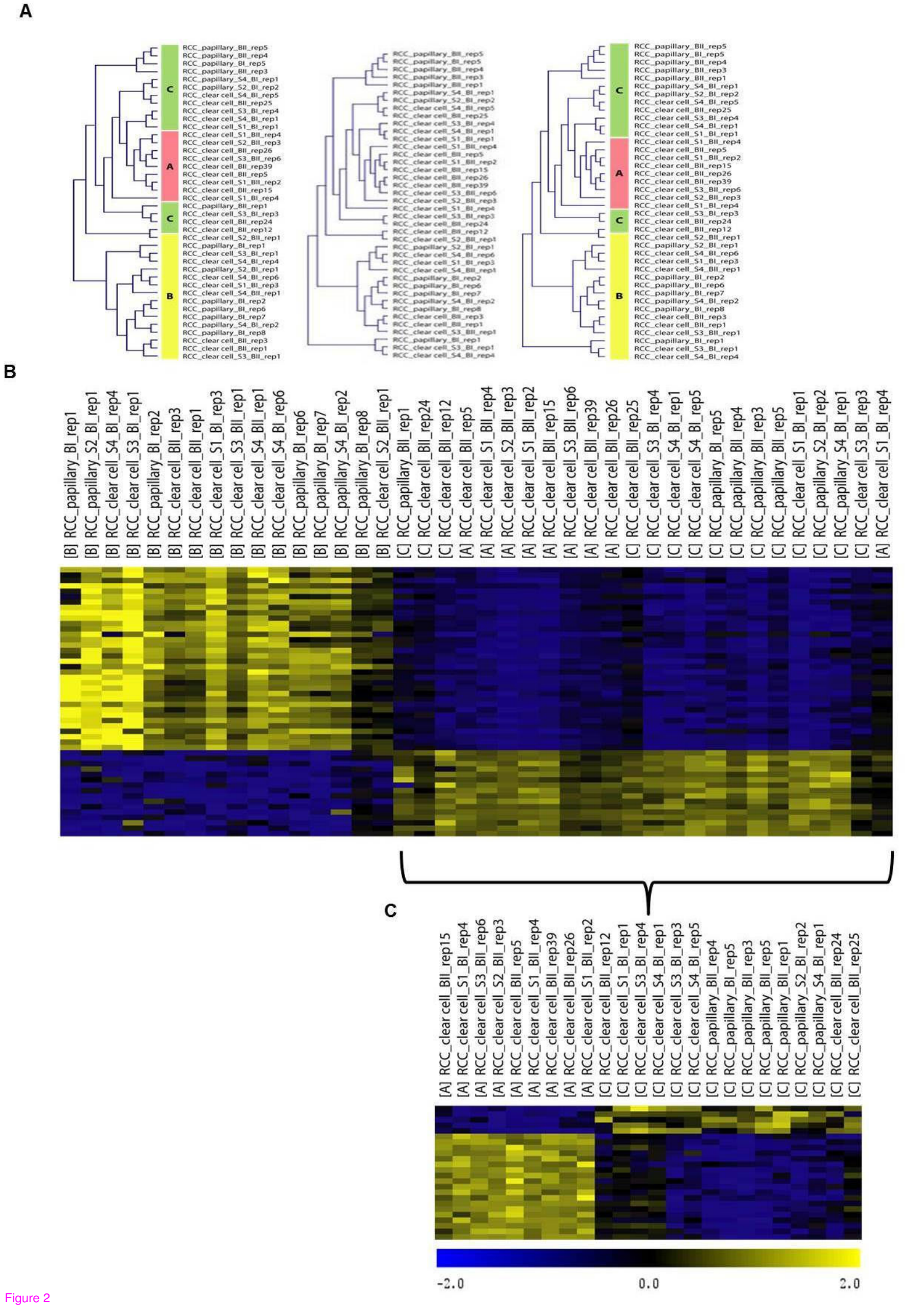
A
C
A
C
B
RCC_papillary_BII_rep5
RCC_papillary_BII_rep4
RCC_papillary_BI_rep5
RCC_papillary_BII_rep3
RCC_papillary_S4_BI_rep1
RCC_papillary_S2_BI_rep2
RCC_clear cell_S4_BI_rep5
RCC_clear cell_BII_rep25
RCC_clear cell_S3_BI_rep4
RCC_clear cell_S4_BI_rep1
RCC_clear cell_S1_BI_rep1
RCC_clear cell_S1_BII_rep4
RCC_clear cell_S2_BII_rep3
RCC_clear cell_BII_rep26
RCC_clear cell_S3_BII_rep6
RCC_clear cell_BII_rep39
RCC_clear cell_BII_rep5
RCC_clear cell_S1_BII_rep2
RCC_clear cell_BII_rep15
RCC_clear cell_S1_BI_rep4
RCC_papillary_BII_rep1
RCC_clear cell_S3_BI_rep3
RCC_clear cell_BII_rep24
RCC_clear cell_BII_rep12
RCC_clear cell_S2_BII_rep1
RCC_papillary_BI_rep1
RCC_clear cell_S3_BI_rep1
RCC_clear cell_S4_BI_rep4
RCC_papillary_S2_BI_rep1
RCC_clear cell_S4_BI_rep6
RCC_clear cell_S1_BI_rep3
RCC_clear cell_S4_BII_rep1
RCC_papillary_BI_rep2
RCC_papillary_BI_rep6
RCC_papillary_BI_rep7
RCC_papillary_S4_BI_rep2
RCC_papillary_BI_rep8
RCC_clear cell_BII_rep3
RCC_clear cell_BII_rep1
RCC_clear cell_S3_BII_rep1
B
[B] RCC_papillary_BI_rep1
[B] RCC_papillary_S2_BI_rep1
[B] RCC_clear cell_S4_BI_rep4
[B] RCC_clear cell_S3_BI_rep1
[B] RCC_papillary_BI_rep2
[B] RCC_clear cell_BII_rep3
[B] RCC_clear cell_BII_rep1
[B] RCC_clear cell_S1_BI_rep3
[B] RCC_clear cell_S3_BII_rep1
[B] RCC_clear cell_S4_BII_rep1
[B] RCC_clear cell_S4_BI_rep6
[B] RCC_papillary_BI_rep6
[B] RCC_papillary_BI_rep7
[B] RCC_papillary_S4_BI_rep2
[B] RCC_papillary_BI_rep8
[B] RCC_clear cell_S2_BII_rep1
[C] RCC_papillary_BII_rep1
[C] RCC_clear cell_BII_rep24
[C] RCC_clear cell_BII_rep12
[A] RCC_clear cell_BII_rep5
[A] RCC_clear cell_S1_BII_rep4
[A] RCC_clear cell_S2_BII_rep3
[A] RCC_clear cell_S1_BII_rep2
[A] RCC_clear cell_BII_rep15
[A] RCC_clear cell_S3_BII_rep6
[A] RCC_clear cell_BII_rep39
[A] RCC_clear cell_BII_rep26
[C] RCC_clear cell_BII_rep25
[C] RCC_clear cell_S3_BI_rep4
[C] RCC_clear cell_S4_BI_rep1
[C] RCC_clear cell_S4_BI_rep5
[C] RCC_papillary_BI_rep5
[C] RCC_papillary_BII_rep4
[C] RCC_papillary_BII_rep3
[C] RCC_papillary_BII_rep5
[C] RCC_clear cell_S1_BI_rep1
[C] RCC_papillary_S2_BI_rep2
[C] RCC_papillary_S4_BI_rep1
[C] RCC_clear cell_S3_BI_rep3
[A] RCC_clear cell_S1_BI_rep4
C
[A] RCC_clear cell_BII_rep15
[A] RCC_clear cell_S1_BI_rep4
[A] RCC_clear cell_S3_BII_rep6
[A] RCC_clear cell_S2_BII_rep3
[A] RCC_clear cell_BII_rep5
[A] RCC_clear cell_S1_BII_rep4
[A] RCC_clear cell_BII_rep39
[A] RCC_clear cell_BII_rep26
[C] RCC_clear cell_BII_rep12
[C] RCC_clear cell_S1_BI_rep1
[C] RCC_clear cell_S3_BI_rep4
[C] RCC_clear cell_S4_BI_rep1
[C] RCC_clear cell_S3_BI_rep3
[C] RCC_clear cell_S4_BI_rep5
[C] RCC_papillary_BII_rep4
[C] RCC_papillary_BI_rep5
[C] RCC_papillary_BII_rep3
[C] RCC_papillary_BII_rep5
[C] RCC_papillary_BII_rep1
[C] RCC_papillary_S2_BI_rep2
[C] RCC_papillary_S4_BI_rep1
[C] RCC_clear cell_BII_rep24
[C] RCC_clear cell_BII_rep25
-2.0
0.0
2.0

Figure 2

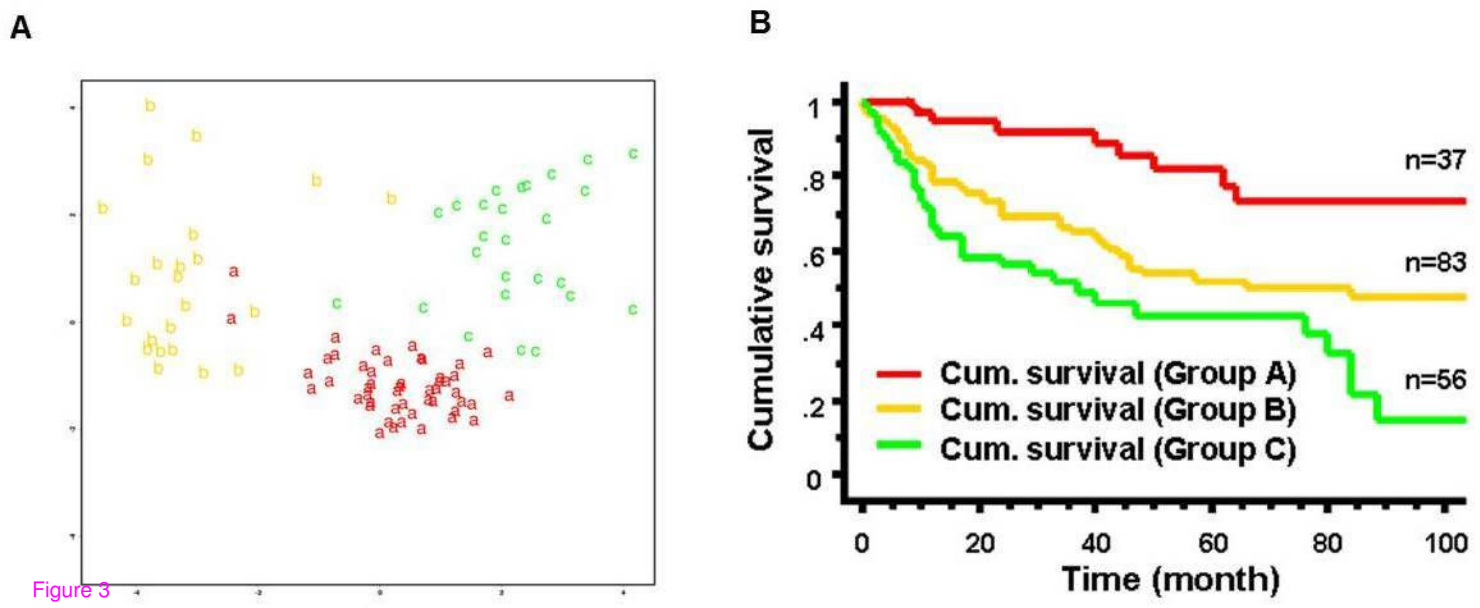
A
B
Cumulative survival
1
.8
.6
.4
.2
0
n=37
n=83
n=56
Cum. survival (Group A)
Cum. survival (Group B)
Cum. survival (Group C)
0
20
40
60
80
100
Time (month)

Figure 3

Additional files provided with this submission:

Additional file 1: 5503591366166554_add1.xls, 45K
http://www.biomedcentral.com/imedia/1291747079769895/supp1.xls
Additional file 2: 5503591366166554_add2.pdf, 1942K
http://www.biomedcentral.com/imedia/2522934207698953/supp2.pdf
Additional file 3: 5503591366166554_add3.xls, 31K
http://www.biomedcentral.com/imedia/2077056340769895/supp3.xls
Additional file 4: 5503591366166554_add4.pdf, 118K
http://www.biomedcentral.com/imedia/1431064588769895/supp4.pdf
Additional file 5: 5503591366166554_add5.xls, 71K
http://www.biomedcentral.com/imedia/9615172676989537/supp5.xls
Additional file 6: 5503591366166554_add6.xls, 32K
http://www.biomedcentral.com/imedia/1938986459769895/supp6.xls
Additional file 7: 5503591366166554_add7.pdf, 140K
http://www.biomedcentral.com/imedia/1170980095769895/supp7.pdf
Additional file 8: 5503591366166554_add8.xls, 31K
http://www.biomedcentral.com/imedia/1082684034769895/supp8.xls
Additional file 9: 5503591366166554_add9.xls, 247K
http://www.biomedcentral.com/imedia/6134328497698953/supp9.xls
Additional file 10: 5503591366166554_add10.xls, 28K
http://www.biomedcentral.com/imedia/7502526676989536/supp10.xls
Additional file 11: 5503591366166554_add11.pdf, 1387K
http://www.biomedcentral.com/imedia/1084437217698953/supp11.pdf
Additional file 12: 5503591366166554_add12.xls, 30K
http://www.biomedcentral.com/imedia/1894469720769895/supp12.xls